General education vs vocational training: How do they affect individual labour market performance?

Pavlina Karasiotou

Abstract

Using the Panel Study of Belgian Households (PSBH, waves 4 to 10) we estimate the effects of education (initial and life-long, general and vocational) on incomes, labour supply and unemployment. This allows for a decomposition of the economic returns of education on earnings in two parts; one attributed on wages and one on employment time. The sample includes individuals 18-65 who have completed initial education at school or university, who are at the labour market and receive income from working. Individuals who are currently following a vocational after-school course are also included in the sample. We use Hausman- Taylor estimators, which are consistent in the presence of correlation between the unobserved individual effects and the explanatory variables and at the same time produce estimates for the time- invariant variables. The results show a large positive effect of initial education both on earnings and on employment; however, this positive effect is smaller for those who follow vocational or technical education while at school. On the other hand, continuous vocational training and life-long learning result to extra gains on top of the gains from initial education. There is a similar picture in the effect of education on unemployment time and labour supply.

JEL classification: J24, J31
Keywords: Mincer regressions, panel data models, general education, vocational training, earnings, employment, returns to education

1. Introduction

During the Lisbon European Council in March 2000, the European Union (EU) set as one of its basic goals to become the "most competitive and dynamic knowledge-based society in the world" ([1]), while the Barcelona European Council, in March 2002 ([2]), reaffirmed this important role. To this end, the development of general and vocational education throughout Europe, the harmonisation of education systems (with the Bologna process in tertiary education for example), the introduction of new learning systems such as distance learning, the development of lifelong learning and a number of other issues are intermediate goals as well tools to the accomplishment of the final goal.

Education, general and vocational, initial and life-long is obviously a goal and a tool for economic development and growth at least from the point of view of policy makers. However, this does not explain why the individual would like to be involved in such a procedure of continuous learning, which obviously entails a very high cost (in money, time, effort). It does not explain what are the individual gains, which would induce individuals to participate; and moreover it does not answer the question as to which kind of education should individuals engage in. When policy makers talk about gains they are mostly interested in macroeconomic goals such as GDP growth and unemployment rate, implying that in such way also microeconomic goals would be achieved (e.g. the wages of each individual would increase or

([1]) Lisbon European Council, March 2000. Available from Internet: http://www.europarl.eu.int/summits/lis1_en.htm#a [Cited 10.11.2003]
([2]) Barcelona European Council, March 2002. Available from Internet: http://europa.eu.int/comm/barcelona_council/index_en.html [Cited 10.11.2003]

that the probability and duration of unemployment would decrease for all). This however is not always obvious nor an immediate result.

We want to show here what are the individual gains from education and shed some light in the dispute between vocational and general education advocates. The international literature shows that more educated individuals earn more, but we want to measure these gains for Belgium, the country of interest in the paper; moreover we will try to show which kind of education generates this increase, i.e. is general education more profitable than vocational/ technical education, is the initial education better than after school lifelong learning, etc.

We follow and expand the traditional human capital model developed by Becker (1964) and estimate the earnings equation developed by Mincer (1974), which is the most frequently used model in the estimation of returns to education. However our model differs from the traditional Mincerian earnings equation in a number of points; first instead of using only years of schooling as a measure of education we distinguish between years of general and years of vocational education while at school and we also introduce vocational life-long education and training. In such way we get a clearer idea about the effect of different kinds of education on earnings. Secondly we go further by estimating separately the effect of education on monthly wages, on labour supply and on unemployment. The sum of these three separate effects is the total effect of education on individual annual earnings. This last characteristic of our model is an extension of the human capital model, since it implies that individual gains from education come not only from increased wages but also from better employment perspectives. Our findings suggest that the gains in terms of time in the labour market and employment are significant for better-educated individuals. Thirdly we introduce a number of other explanatory variables such as occupation, sector of activity, family background etc., in comparison with the traditional Mincer equation that only controls for schooling years and experience.

The second basic element of this paper is that it attempts to control for unobserved individual effects, such as ability, personal ambitions etc, which when correlated with the explanatory variables (here with education) bias the results. The data we use here are from the Panel Study of Belgian Households (PSBH), a longitudinal database that includes information on individual income, education, labour market experience, demographics, etc. The two traditional panel data estimators, the within and the random effects, however are inadequate for estimation here, each for different reasons. The random effects estimator produces inconsistent and biased estimates in the case of correlation between the unobserved individual effects and the explanatory variables. On the other hand, the within estimator (otherwise called the fixed effects estimator) always produces consistent estimates, even in the presence of correlation; however, all time-invariant variables are eliminated by the within transformation (deviation from individual means). Here the main variable under examination, education, is a time-invariant variable, and the within estimator is basically of no use.

An answer to both these problems (time-invariant variables and correlation) is the Hausman-Taylor estimator. It divides explanatory variables into four groups, depending on their time variability and their relation to the unobserved individual effects: time-invariant exogenous variables, time variant exogenous variables, time variant endogenous and time- variant endogenous variables. It then uses the deviations from the means (within estimators) of the exogenous variables to instrument endogenous time invariant variables. One positive aspect of the HT estimator is that the exogeneity assumption of the variables does not need to be based on a priori assumptions, but rather can be tested with the Hausman specification test.

Our results show a large positive effect of initial education both on earnings and on employment; however, this positive effect is smaller for those who follow secondary vocational or technical education while at school. This indicates the importance of developing general skills while at school such as mathematical, analytical etc and acquiring general knowledge. On the other hand, continuing vocational training and life-long learning generally result to extra gains on top of the gains from initial education; however these gains depend on the initial education a person has received. There is a similar picture in the effect of education on unemployment time and labour supply, but the majority of the estimates are not statistically different from zero.

In the following section 2 we briefly present some basic characteristics of the Belgian education system and discuss findings in the literature on the relationship between general and vocational education. In Part 3 we present the connection between annual income, monthly earnings, labour supply and unemployment. In Part 4 we develop the econometric model we are using in the estimation and in part 5 we present and discuss in detail the sample and the construction of the variables. Part 6 presents and discusses the results. Part 7 concludes.

2. The multiple dimensions of education

a. The Belgian system of general and vocational education

Belgium is a federal State with multiple levels of power resulting in a division of the way education is administered and financed. There exist three linguistic communities (Flemish-speaking, French speaking and German-speaking), three geographical regions (Flemish, Walloon and Brussels) and five levels of power: (the national, the community, the region, the province and the municipality). The structure of the education system is not really different between the communities. For example, the duration of primary, secondary and tertiary education is the same, minimum school leaving age is the same in all regions, secondary education is divided in general, artistic, technical and vocational throughout the country etc. These characteristics make the educational system in general comparable between the communities. However, there exist differences in the administration and financing that lead many to argue that the results and effects of education and especially vocational and lifelong education will differ substantially among the communities. We will examine that issue further in the following.

Initial education is divided into three main levels:

a) Basic education that includes pre-school and primary education. Primary education lasts 6 years from the age of 6 to 12.

b) Secondary education, from the age of 12 to 18. Up to 1971, secondary education was organised into two three-year cycles (lower and upper secondary education). In 1971 secondary education was reformed and is now organised into three two-year cycles (a *common* or *foundation* cycle which is common to all students, an *orientation* cycle and a *determination* cycle). There is also a fourth cycle organised in full-time advanced vocational education in several schools for certain training streams, e.g. medical and psychiatric hospital nursing. This is a 3-year programme, sometimes preceded by a preparatory year.

The new 1971 system was gradually adopted by all school types in all three communities. The main purposes of the reform were to postpone up to the age of 15/16, pupils' choices about continuing to tertiary education, or follow vocational training and keep pupils one further year to secondary education. The first two cycles of the reformed system correspond to the former *lower* secondary education and the third cycle to the former *upper* secondary education. There exist four different forms of secondary education: general, artistic, technical and vocational. The selection of a secondary education stream usually defines at least up to a point, the post-school and career choices of pupils. After the age of 15/16 pupils can also enter part-time secondary education.

c) Tertiary education is organised in *short type non-university education*, *long type non-university education* and *university education*. Until 1990 several short type education courses were organised as two year courses; however due to changes in the European union requirements for recognition of certificates, all short type courses are now of duration of at least three years. Short type education offers mainly practical knowledge and skills and prepares students for working life. Long type non-university education is organised in 4- or 5-year courses and leads to advanced scientific and/or technological qualifications. They award *candidatures* and *licences* in the same way as universities.

Finally, university education has three cycles: i) first cycle studies *(candidature)*, which lasts two to three years and opens the door to the second cycle of university studies, ii) second cycle studies *(licence)*, which lasts two to three years and ends with the submission of a dissertation and iii) third cycle studies which can last several years. This last cycle includes advanced, specialised studies and research and leads to various degrees such as a Doctorate (a PhD in the Anglo- Saxon system) or a DEA (diplôme des études approfondies) usually equivalent to a master’s degree in the Anglo-Saxon system.

Apart from the vocational- technical stream of secondary education and the short type non-university education, vocational adult education and training is provided by a number of agents and institutions and targets to young people who left school before completing secondary education, to young adults who completed secondary education but did not continue to tertiary education, to employed adults who want to increase their skills and keep up with new technologies and labour market conditions, unemployed, SME (small and medium sized enterprises) owners, civil servants etc.

Young people who leave school at the age of 15/16 can enter alternance education which includes a part of theoretical teaching in class and a part of working in a firm. The “alternance training” population has been enlarged to include young adults up to 25 years.

Continuing vocational training includes initiatives by the employers, the authorities, social advancement education and individuals. *Social advancement education* is the main adult vocational-training stream in Belgium. It promotes improved vocational, social, cultural, and scholastic integration; it addresses the training needs and requests of companies, administrations, the education system, and, in a general way, of socio-economic and cultural environments (Eurydice, 2001). Education for social advancement is organised at the secondary and tertiary education levels.

Several public institutions are involved in training and employment services as well as accreditation. The main institutions are FOREM for the Walloon region, VDAB for the Flemish region, Bruxelles-Formation for Brussels and the Arbeitsamt der Deutschsprachingen Gemeinschaft, for the german-speaking community.

The Belgian education system is rather complicated both because of the special structure of the Belgian state, but also because of the existence of multiple education streams. Individuals have several options and can choose from a variety of types and duration of courses, which is the norm for most of education systems in the European Union. The multiplicity of the educational system poses some difficulties in measuring time in education and must be taken into account when estimating returns to education.

b. The different types of education in the literature

Given the multiple streams of education, one may wonder what is the evidence in the literature in favour and against each stream. Is it preferable for individuals to follow general or vocational education while at school? Should individuals continue with tertiary and university education or follow some vocational training after school? And is lifelong education of any use to individuals who have already acquired quite high initial education?

Mane (1999) argues that from those pupils who do not continue to tertiary education, those with vocational courses in their high-school curriculum do considerably better than those with only general skill courses. He analyses the effect of the number of vocational and general courses in high school on the number of months employed, on hourly wage rate and on yearly earnings; in that way his analysis approximates the analysis done here. The positive effects of vocational courses for pupils who do not continue to college are considerable, both in terms of working time and wages, but seem to diminish with time in the labour market. However this analysis applies for the U.S. education system where post-secondary non-university education is rather limited and individuals have fewer opportunities to acquire vocational qualifications and training after school, compared to students in Europe. On the other hand, European surveys report extra gains for individuals with academic qualifications compared to individuals with the same years of schooling but with vocational qualifications (Dearden et al, 2000, McInstosh, 2004, Conlon 2001).

Stern et al. (1997) find using US data, that when high-school and two-year college students work in a school supervised job (something close to the Belgian "alternance system" and the German "dual system") they do better at school, they develop a better connection between school and work and do better (although this effect diminishes) in the labour market compared to those who work in an non-school supervised job or do not work while at school.

In a paper on the returns to education in the US sub-baccalaureate labour market, Grubb (1997) notes that the economic benefits to non-college post-secondary education vary substantially depending on several factors. The benefits are higher for individuals who have received a significant amount of post-secondary education (more than one year) and who have completed a certificate; the benefits also depend on the field of studies followed as well as on whether subsequent employment is related to studies. These findings are also confirmed for the UK. Dearden et al (2000) report that academic qualifications are more profitable for those who acquire a skilled job, while McIntosh (2004) reports differences in vocational and academic premia depending on whether the individual works in the private or public sector. Furthermore they both find that there exist gender differences; men and women with vocational qualifications do not do equally well in all sectors. Men receive the largest returns for professional qualifications such as accountancy, while for women the vocational qualifications with the highest returns are nursing and teaching.

It seems that the returns to vocational and general education also depend on the form of the education system. Cooke (2003) discusses that during the 1980's when the German education system was highly stratified with pupils not being able to move easily from one education stream to another, post-secondary vocational education did not have any significant economic effect. On the contrary, during the 1990's the education system became more flexible and students turned more towards general education; then graduates with after school vocational education earned much more than graduates with no vocational education. Some further findings are that apprenticeship in a firm (dual system) is more effective than other vocational education types and secondly that the effects of vocational education become obsolete relatively quick, providing in that way an argument for lifelong learning.

In general there exists a big debate on whether general/ academic education is better than vocational/ technical education and training. The findings are highly dependent upon the educational system, the specific assumptions and questions in each survey and upon the datasets, which very often provide very limited information on all the qualifications an individual has acquired during his academic and professional life. Here we will use the information provided by PSBH to assess differences between general and vocational, initial and life-long education for Belgium and compare these findings to similar findings for other countries.

3. Monthly or annual earnings? A decomposition story

When we want to estimate the returns of education on individual incomes, does it make any difference whether we choose annual, monthly, weekly or hourly earnings? Are the results affected and does this difference have any economic interest and meaning? In the literature measuring the returns of education on earnings, these questions are not often discussed. The choice of time frame over which to measure earnings is usually dictated by necessity; different datasets include information on different earnings' measures (Card, 1999).

However, the returns of schooling on annual earnings reported in the literature are usually higher than the returns on monthly or hourly earnings (see for example Card (1999) and Mincer (1974)). A natural question that rises is why this happens. The explanation is relatively easy to give. Just think how we arrive to calculate annual income:

annual income=(hourly wages)×(hours worked per day)×(days worked per month)×(months worked per year).

There is therefore a time element that we often forget: individual earnings are not only the amount of money a person receives; they are also the time a person spends working. If we want to go a step further we should see that an individual receives income a) when he is in the labour market i.e. is not in education or retirement or military service etc, and b) when he is not unemployed.

But is employment time affected by education? If the answer to this question was negative, education returns should be the same independently of which time measurement of income we used in the calculations. As said before, this is not the case, so at first thought education does affect employment time and so, when we calculate education returns on earnings, time worked is an element that should enter explicitly into our problem.

The education effect on working time is usually positive, in other words, more educated individuals work more, so this explains why returns on annual earnings are higher than returns on monthly (hourly, etc) earnings.

On the other hand, the next question should be why the education/ working time relationship is positive. This may occur, either because labour supply increases with schooling level or because more educated individuals become less frequently unemployed and/or spent less time in unemployment. Both of these explanations are plausible. Table 1 reports recent OECD data (2003) for Belgian males and females concerning labour force participation (LFS) and unemployment by education level. Education levels are slightly different determined in OECD data compared to the data we will use in the estimations, but the comparison is relatively easy.

Table 1 Labour force participation and unemployment rates by educational attainment, Belgium

	Below upper secondary education (<12yrs)	*Upper secondary & post-secondary non-tertiary education (12-14 yrs)*	*Tertiary- type B (non-universitary) education (13-14 yrs)*	*Tertiary-type A (univ.) & advanced research programmes(≥18yrs)*	*All levels of education*

	Labour force participation rates (2001) by education level and gender for 25 to 64 year-olds				
Males	69	87	91	92	81
Females	39	69	81	84	60
	Unemployment rates (2001) by education level and gender for 25 to 64 year-olds				
Males	7.4	4.4	2.5	2.5	4.8
Females	10.4	6.9	2.4	3.8	6.3

Source: OECD, (2003) Extracts from Tables A12.1 and A12.2

The differences both in unemployment rates and labour force participation rates by level of education are quite high for both men and women. For example, a difference of at least six years in education (less than upper secondary education to university education) implies an increase in labour force participation of 33% (around 5.5 percentage points per year) and a decrease in unemployment 66% (11 percentage points per year) for males. The differences are not so big once we compare university education with other tertiary education types or with upper secondary education, but they are still important.

The reasons why education improves labour market experiences of individuals are explored quite extensively in the literature. Technological changes (Bartel and Lichtenberg 1987, Nickel and Bell 1991), screening from employers (Holzer, 1996), more efficient on- and off-the job searching (Mincer, 1991), psychological and sociological reasons (discouragement and exit from labour force for less-skilled workers, family decisions such as the raising of children) are only few of the reasons that have been explored as explanations for why education improves employment opportunities and labour force participation rates.

Irrespectively of the reasons behind the better labour performance for better educated individuals we are interested here to estimate the gap between returns to education for annual and monthly earnings and see which part of this gap can be attributed to higher labour force participation and which part to less unemployment. As far as we know these estimations are not available up to now for Belgium.

We use the model developed by Ashenfelter and Ham (1979) (from now on referred as A&H). A&H assume that labour supply h* is the sum of employed (h) and unemployed (u) months h*=u+h. All months spent in activities other than in a job or in an active search of a job (i.e. unemployment) are assumed as a part of an individual's decision not to offer his labour into the market. This is true to the extent that no compulsory schooling laws and other restrictions (e.g. compulsory military service) influence a persons' available time for work or job search. With a monthly wage w, *desired* earnings are then:

$$wh^* = w(h+u) \Rightarrow$$

$$wh^* = wh(1+u/h)$$

where wh are annual realised earnings from work and u/h is the fraction of unemployment to employment months. By taking logs the relationship becomes:

$$\ln wh = \ln w + \ln h^* - \ln(1+u/h)(^3) \qquad (2.1)$$

Equation (2.1) is a decomposition of annual earnings into wages, labour supply and unemployment/ employment ratio. However it works only in cases where income from work exists, i.e. it excludes individuals with h=0 and u=12, who receive some replacement income (unemployment benefits, pre-pension benefits etc). It is a common practice in the literature to

[3] A&H actually use the relationship $\ln wh = \ln w + \ln h^* - (u/h)$ instead of (1). The simplification *ln(1+u/h)≈ u/h* that they make is true for small values of *u/h*, namely for *u/h*<0,1. This is a plausible assumption for 1979 when the A&H paper was written and one can assume that individuals spent only a small fraction of their time in unemployment. However today's experience saws that we can have u/h ratios very much above 0,1 (for example an individual with 10 months of unemployment and 2 months of employment has a u/h ratio equal to 5). This implies that the simplification of A&H may distort our results and consequently we will not use it in our estimations.

include in the estimations only those who receive wages; however as we will comment further below, such exclusions probably overestimate the effect of education on earnings (the well known *selection bias* problem).

By differentiating (2.1) an immediate result is:

$$\partial \ln(wh)/\partial S = \partial \ln w/\partial S + \partial \ln h^*/\partial S - \partial \ln(1+u/h)/\partial S \qquad (2.2)$$

where S is a measure of schooling, usually years.

Equation (2.2) decomposes the effect of schooling into its parts, i.e. the effect of schooling on wages, labour supply and u/h. Assuming that annual and monthly income, labour supply and unemployment are all linear functions of schooling and a vector of other explanatory variables as equation (2.3) suggests, then all we need is a method that produces consistent estimates of the β's and the γ's so as to estimate the gap between education returns on annual and monthly earnings and determine its source.

$$\ln y_i = \alpha_i + \beta_i S + x_i'\gamma + \varepsilon_i, \qquad (2.3)$$

where *lny* represents *ln(wh), lnw, ln(1+u/h)([4]) and lnh** respectively and x is a vector of explanatory variables.

Any consistent estimates of the return to schooling (but also of the other variables) in (2.3) will satisfy the relationship

$$\hat{\beta}_{wh} = \hat{\beta}_w + \hat{\beta}_{h*} - \hat{\beta}_{u/h} \qquad (2.4)$$

which is equivalent to (2.2). Then we have four possible cases that explain the difference between monthly and annual earnings. The first is the trivial one that $\hat{\beta}_{wh} - \hat{\beta}_w = 0$ implying that education has no effect on time spent on employment and we have already mentioned that the literature concludes otherwise. Secondly $\hat{\beta}_{wh} - \hat{\beta}_w = \hat{\beta}_{h*}$ would imply that education does not affect unemployment experiences of individuals and all the difference between annual and monthly earnings is due to the effect that education has on labour supply. The third case is the opposite one namely $\hat{\beta}_{wh} - \hat{\beta}_w = -\hat{\beta}_{u/h}$, implying that labour supply is independent of schooling and the last case is that all components of equation (2.4) are different from 0 and so the difference between annual and monthly earnings can be explained from the effect of education on labour supply as well as unemployment.

4. The Model

We are using here longitudinal data, which, compared to cross-section data, allow for control for omitted or unobservable individual-specific effects in the specification of an econometric model.

The two traditional panel data estimators are the within (fixed effects) and random effects estimator. In the within estimator, data are transformed into deviations from individual means and OLS is performed in the transformed data. This method always produces consistent estimators irrespectively of potential correlation of the unobserved individual effects with the explanatory variables. However the within estimator has a major defect: all time-invariant variables are eliminated by the transformation. This problem is of great importance here since the main variable of interest (schooling) is time invariant.

On the other hand the random effects estimator is a weighted average of the within and between (data are transformed into individuals means) estimators. This implies that the

[4] From now on, instead of (1+u/h) we will use u/h for notational convenience

random effect estimator is inconsistent in case of correlation between the explanatory variables and the unobserved individual effect. The Hausman specification test, allows for testing for correlation by comparing within and random effects estimators. Correlation between the effects and the variables are a common phenomenon in the estimation of education returns; schooling can be correlated with individual ability and ambition (unobserved effects).

As an alternative to these, Hausman and Taylor (1981) (thereafter HT) have proposed an instrumental variable estimator that has neither of the defects of the within estimator and at the same time is consistent. Under their assumptions, their estimator is an efficient GMM estimator. Let us consider a specific model:

$$y_{it} = X_{it}\beta + Z_i\gamma + a_i + u_{it} \tag{5.1}$$

where i=1…N, and t=1,…T. The X_{it} are individual time-variant variables and Z_i are individual time- invariant variables. a_i is IID$(0,\sigma_\alpha^2)$ and may be correlated with parts of X and Z. u_{it} is IID$(0,\sigma_u^2)$ and is assumed to be uncorrelated with both the explanatory variables and the effects. The two main assumptions in the HT model are firstly that only the time-invariant component of the error term is correlated with explanatory variables and secondly that not all of the explanatory variables are correlated with a_i.

HT split $X = [X_1, X_2]$ and $Z = [Z_1, Z_2]$, with X_1, Z_1 assumed exogenous and not correlated with a_i and u_{it}, while X_2 and Z_2 are uncorrelated with u_{it} but correlated with a_i. X_1 is $NT \times k_1$, X_2 is $NT \times k_2$, Z_1 is $NT \times g_1$ and Z_2 is $NT \times g_2$. The main idea then behind the HT estimator is that the exogenous time-variant within estimators (X_1) are used as instruments for estimating time-invariant endogenous coefficients (Z_2). Deviations from individual means are used as instruments for X_1 and X_2 (since they are time-variant) and Z_1 are used as their own instruments (since they are exogenous).

We use the following notation for projections. For any matrix A, its projection onto the column space of A is defined as: $P_A = A(A'A)^{-1}A'$ and $Q_A = I - P_A$ is defined as the projection onto the null space of A. More intuitively P_A, transforms a vector of variables into individual means and Q_A into deviations from means. Finally, we define V a NT×N matrix of individual-specific dummy variables. The model in (5.1) can be written as:

$$y = X\beta + Z\gamma + V\alpha + u \tag{5.2}$$

The HT partition of the variables into the four groups described below and their exogeneity assumptions imply the following orthogonality conditions:

$$E\left[X_{1it}'(\bar{y}_i - \bar{X}_{it}'\beta - Z_i'\gamma)\right] = 0 \tag{OC1}$$

$$E\left[Z_{1it}'(\bar{y}_i - \bar{X}_{it}'\beta - Z_i'\gamma)\right] = 0 \tag{OC2}$$

$$E\left[(y_i^* - X_i^*\beta) \otimes X_{it}\right] = 0 \qquad \text{(OC3)},$$

where the asterisks imply that the variables are expressed as deviations from means. These orthogonality conditions along with the disturbance covariance matrix Ω, which is analysed below, are used in a method of moments procedure.

The disturbance covariance matrix $\Omega \equiv \mathrm{cov}(V\alpha + u) = \sigma_u^2 I_{TN} + T\sigma_\alpha^2 P_V$ is a block-diagonal matrix that has the random effects structure. HT show how we can obtain consistent estimates of $\sigma_u^2$and σ_α^2.

Another assumption for efficient estimation is the system homoskedasticity assumption $E(u_i u_i^{'}/Z_i) = E(u_i u_i^{'})$. For it to hold we use the non-singular matrix $\Omega^{-1/2} = Q_V + \theta P_V$, (where $\theta^2 = \sigma_u^2(\sigma_u^2 + T\sigma_a^2)^{-1}$ ([5])) to transform the error covariance matrix into a scalar matrix:

$\Omega^{-1/2}\Omega\Omega^{-1/2} = \sigma_u^2 I$.

Transforming equation (5.2) by $\Omega^{-1/2}$ is equivalent to a differencing of the observations:

$$\Omega^{-1/2} y = \Omega^{-1/2} X\beta + \Omega^{-1/2} Z\gamma + \Omega^{-1/2}(V\alpha + u) \quad ([6]) \qquad (5.3)$$

Then with an instrument set I, based on the partition of the explanatory variables into endogenous and exogenous, 2SLS is performed on (5.3) taking X_1 and Z_1 as exogenous.

This yield estimators of the form:

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \left[(X,Z)'\Omega^{-1/2} P_A \Omega^{-1/2}(X,Z)\right]^{-1}(X,Z)'\Omega^{-1/2} P_A \Omega^{-1/2} y \qquad (5.4).$$

The HT estimator uses as instrument set:

$$I = (Q_V X_1, Q_V X_2, P_V X_1, Z) \qquad (5.5).$$

The order condition for the HT to exist is $k_1 \geq g_2$ (i.e. the exogenous time-variant variables to be at least as many as endogenous time-invariant). (Cornwell & Rupert, 1988). In case where $k_1 = g_2$, the model is just identified and the HT estimator is the within estimator for the time-variant variables, while when $k_1 > g_2$, the model is over-identified and the HT estimator is more efficient than the within estimator.

Compared to conventional IV estimators, in the HT estimator the k_1 exogenous time-variant variables, which are the only candidates for identifying instruments are included in the structural equation (5.2). As HT point in their work, this works because only the time invariant component of the disturbance is correlated with X_2 and Z_2.

The exogeneity of variables and subsequently their use as instruments to compute the HT estimator is tested by the well-known Hausman specification test. It is based in the fact that β is always identified and $\hat{\beta}_{Within}$ provides a consistent benchmark against which the exogeneity restrictions in the model can be tested by comparing $\hat{\beta}_{Within}$ with $\hat{\beta}_{HT}$. The null hypothesis is of the form:

$$H_0 : p\lim_{N\to\infty}\frac{1}{N}X_{1i.}^{'}a_i = 0 \text{ and } p\lim_{N\to\infty}\frac{1}{N}Z_{1i.}^{'}a_i = 0.$$

Acceptance of the null hypothesis implies that the X_1, Z_1 variables are truly exogenous and therefore can be used as legitimate instruments.

After the HT paper, Anemiya & MaCurdy (1986) (AM) and Breusch, Mizon & Schmidt (1989) (BMS) have proposed improved versions of the HT estimator. More specifically they

---

([5]) For unbalanced panels θ is expressed as $\theta_i^2 = \sigma_u^2(\sigma_u^2 + T_i\sigma_a^2)^{-1}$, depending on the individual number of observations (Gardner, 1998)

([6]) More precisely, we are dealing here with a system consisting of a single structural equation and two multivariate reduced form equations:

$$\Omega^{-1/2}Y_{it} = \Omega^{-1/2}X_{it}\beta + \Omega^{-1/2}Z_i\gamma + \Omega^{-1/2}u_{it}$$
$$X_{2it} = X_{1it}\pi_{11} + Z_{1i}\pi_{12} + Q_V\pi_{13} + \varepsilon_{1it}$$
$$Z_{2i} = X_{1it}\pi_{21} + Z_{1i}\pi_{22} + Q_V\pi_{23} + \varepsilon_{2it}$$

have proposed different sets of instrumental variables and they prove that with legitimate instruments their estimators are at least as efficient as the HT. However, because of the stricter requirements they place on the instruments and the fact that they can be estimated only for balanced panels, we will continue only with the estimation of the HT. Moreover, as pointed out by Baltagi and Khanti-Akom (1990), the gains in efficiency are rather small when the instruments are legitimate.

We briefly present some studies that have used HT estimator to measure returns to schooling. In their seminal paper Hausman and Taylor (1981) estimate returns to schooling as an application for their estimator. Their instruments are experience, bad health and unemployment last year. The estimated schooling coefficient is 0,125 (s.e.0,0434), which they find to be almost 62% above the within estimate and 30% above the traditional IV estimate.

Cornwell and Rupert use as exogenous variables (instruments) weeks worked, a dummy indicating whether the individual resides in the south of the country, a dummy indicating whether an individual resides in a metropolitan area and marital status. Their estimated HT schooling coefficient is 0,2004 (s.e.0,0783) which is more than double than the random effects estimator for schooling, which is 0,0941 (s.e. 0,017). The Hausman test they performed does not reject the null hypothesis that the effects are uncorrelated with the variables for the HT estimator.

In 1990, Baltagi and Khanti-Akom correct the data by Cornwell and Rupert and their results question the legitimacy of the instruments used. By moving into a new group of instruments, (which includes occupation, living in south, living in metropolitan area and industry) they correct the Hausman test-statistic. However, now the within and HT estimators are not statistically different.

As is apparent from the above and as pointed out in the literature, the selection of legitimate instruments (i.e. time variant exogenous variables) is of vital importance for the validity of the calculated estimators. The exogeneity of variables can be tested with Hausman specification tests, with the within estimators used as benchmarks.

For its apparent virtues compared to the random effects (correction of endogeneity) and the within estimator (estimation of time-invariant coefficients), but also because it suits better our data compared to the AM and BMS estimators (can be applied in the case of unbalanced panels), we will use the HT estimator throughout the paper.

5. Our sample and construction of the variables

To estimate our models we use data from waves 4 to 10 (correspond to years 1995-2001) of the Panel Study on Belgian Households (PSBH). PSBH is a survey conducted jointly from the Universities of Antwerp and Liege and includes longitudinal information on Belgian households and individuals. The survey covers a very broad range of issues including questions on demographics, income, health, unemployment, education etc. Despite the valuable information it contains, PSBH has certainly very important drawbacks that make it difficult to use and extract credible results. More specifically, until 1992 when the European Community Household Panel (ECHP) was introduced the structure of and the questions in the PSBH were constantly changing, making in such way the three first waves not comparable with each other and with subsequent waves. Moreover PSBH is a highly unbalanced database both because individuals drop out or are added in each wave and individuals do not answer all questions. However the estimation methods we are using are robust to unbalancedness and estimations we did only with balanced sample did not produce significantly different results.

From the whole initial PSBH population (waves 4-10) we have excluded individuals less than 18 and more than 65 years old. We have also excluded those individuals who have spent all seven waves in inactivity (retirement, housework or other) or education. Moreover we eliminated from the sample individuals who during the survey were following initial education (primary, secondary or tertiary). In that way years of (initial) education is a time-invariant variable within the sample. However, the sample contains individuals who during the survey period engaged in lifelong education (e.g. vocational courses, language courses, at the workplace or in an educational institution). The effects of initial and continuing education will be separately examined. Vocational after-school training is therefore a time-variant variable within the sample.

Finally, we only use in the estimations individuals who had received some income from work (wages, salaries, paid apprenticeships) during each wave of the survey. This considerably reduced our sample and introduced some further unbalancedness as we exclude waves where individuals are only unemployed (u=12 and h=0).

The final sample used in the estimations includes 10 811 observations, which correspond to 2 765 individuals. Only 426 of those individuals (or 15.41%) are followed throughout all seven periods and another 10% is followed throughout all but one periods.

Very often when we constructed the variables we had to combine information coming from more than one questions, as for example in the case of income variables. In general we constructed and used four major variable categories:

a) The education group, which includes variables on initial (school) education, both general and vocational, and on life-long (after school) vocational education as well as interactions between them.

b) The income group, which includes the two variables on annual income from wages and salaries and on monthly income.

c) The time variables group. This includes the variables measuring months at employment, unemployment and labour supply.

d) Finally we have the group which includes the rest of the explanatory variables. These are either time-variant (marital status, age cohort, health, experience and experience square, sector of employment, profession and region) and time-invariant (sex, nationalty, father and mother education).

The following Table 2 explains each of the variables we are using, while Appendix 1 presents in detail the construction of the variables and the PSBH question they correspond to.

Table 2 Dependent and explanatory variables

Variable name	*Variable description*
education	Years of full time initial (school) education. Tertiary education is also included here
interaction	Interaction between years of initial education and dummy on initial education (=1 if individual has followed the vocational or technical cycle of studies when at school)
voc_educ1	=1 if individual has ever followed a vocational or other course after completing initial education
sgen_voc	=1 if voc_educ=1 & individual has completed general secondary (initial) education
svoc_voc	=1 if voc_educ=1 & individual has completed vocational secondary (initial) education
tshort_voc	=1 if voc_educ=1 & individual has completed short tertiary (initial) education
tlong_voc	=1 if voc_educ=1 & individual has completed long tertiary (initial) education
univ_voc	=1 if voc_educ=1 & individual has completed university (initial) education
voc_school	=1 if the vocational course the individual has followed was in an educational establishment
voc_work	=1 if the vocational course the individual has followed was in the workplace
voc_altern	=1 if the vocational course the individual has followed was both at workplace and at school
voc_forem	=1 if the vocational course the individual has followed was organised by an employment service
voc_train	duration , in weeks, of vocational courses an individual might followed at the survey time
exper	=current age minus the age at which the individual got his/ her first full time and regular job.
exper_sq	square of experience
services	=1 if the individual works at the services sector
professional	=1 if the individual is working as a higher employee, professional, director etc.
BXL	=1 region of residence is Brussels
Flanders	=1 region of residence is Flanders
Wallon	=1 if region of residence is Walloon region
woman	=1 if individual is a woman
nation1	=1 if individual has the Belgian nationality
f_hedu	=1 if father has completed tertiary (university and non-university) education
f_sedu	=1 if father has completed secondary education
f_ledu	=1 if father has no education or has completed primary education
m_hedu	=1 if mother has completed tertiary (university and non-university) education
m_sedu	=1 if mother has completed secondary education
m_ledu	=1 if mother has no education or has completed primary education
ms	=1 if individual is married
hlth	=1 if individual is in good health
age_coh1	=1 for individuals in the age group 18-30
age_coh2	=1 for individuals in the age group 31-45
age_coh3	=1 for individual in the age group 46-65
Dependent variables	
lnwh	logarithm of annual income from wages and salaries
lnw	logarithm of monthly income
lnh*	logarithm of labour supply months (labour supply= employment months+ unemployment months)
lnu/h	logarithm of unemployment to employment months ratio

b. Descriptive statistics

We want to get a first picture of the characteristics of our sample and the relations that exist among variables. Table 3 summarises the income and time variables by years of schooling, while Table 4 summarises explanatory variables.

Table 3: Income and employment by years of education (standard deviation in parentheses).

Education groups	*annual income* ([7])	*monthly income*	*unemployment ratio (u/h)* ([8])	*labour supply (months)*
0-6 yrs	462 092 (173 750)	40 311 (12.983)	0.14 (0.97)	11.72 (1.35)
7-9 or 10 yrs (general)	547 657 (225 302)	46 832 (17.928)	0.07 (0.52)	11.81 (1.14)
7-9 or 10 yrs (vocat.)	487 010 (185 363)	41 983 (13.763)	0.16 (1.04)	11.80 (1.13)
10-12 yrs (general)	587 058 (255 335)	50 628 (21.150)	0.05 (0.50)	11.65 (1.58)
10-12 yrs (vocational)	510 115 (209 827)	44 077 (15.940)	0.07 (0.50)	11.71 (1.42)
13-14 or 15 yrs (short)	612 745 (326 001)	52 105 (26.474)	0.03 (0.40)	11.83 (1.09)
16 yrs (long term)	625 973 (233 103)	53 146 (18.498)	0.04 (0.52)	11.85 (0.96)
14-18 yrs (university)	826 139 (394 224)	69 939 (31.889)	0.05 (0.57)	11.86 (0.93)
All education levels	619 295 (304 725)	52 847 (24.497)	0.06 (0.59)	11.80 (1.16)

The decomposition we are using rules out individuals that are unemployed throughout the whole year, since this implies that either income is 0 or the ratio u/h is not defined, or both. On the other hand we also excluded individuals who are still in education or permanently out of the labour market. These two exclusions led firstly to relatively small u/h ratios, probably underestimating in that way the unemployment effect. Secondly they led to average labour supply close to 12 months, even though less educated individuals exhibit greater variation, implying larger moves in and out of the labour market.

Apparently income, both annual and monthly, rises with education, while general education is also more "profitable" compared to vocational education. Individuals with 6 years or less of education earn 76.2% of the average monthly income for the whole sample, compared to individuals with university education that earn 132% of the average annual income. The income divergence between education groups is even higher when we compare annual instead of monthly earnings. In general it appears from the data description that the differences among education groups are around 1-2 percentage points higher when we compare annual earnings than when we compare monthly earnings, due to the effect of unemployment and labour supply.

As it appears from Table 3 an individual with primary education spends around 4.5 less days per year in the labour market compared to a university graduate, and at the same time has about 9% higher unemployment to employment (u/h) ratio. Again individuals with vocational instead of general secondary education seem less advantaged both in terms of unemployment as well as labour supply. Finally, university graduates have a slightly higher unemployment ratio compared to non-university tertiary education graduates.

In Table 4 we summarise all other variables we are later using in regressions. Their description is in Table 2, while in Appendix 1 we present the PSBH questions they correspond to. With respect to categorical variables we present three different percentages: *overall*, *between* and *within*. The "*overall*" percent corresponds to percentage of observations, *between* corresponds to percentage of individuals and *within* corresponds to the percentage of individuals who have not changed their answer throughout the survey, i.e it is a measure of

([7]) Annual and monthly income are expressed in BEF, in fixed 1995 prices (1 €= 40,34 BEF)

([8]) Employment, unemployment and labour supply are expressed in months the variation of the variable and as a consequence time-invariant variables have a within percent equal to 100. The last column identifies which variables in our sample are time variant (TV) and which are time-invariant (TI)

Table 4: Summary statistics

Name of variable	*Mean (s.d. in parenthesis)*	*Overall (%)*	*Between (%)*	*within percent (%)*	*time variability*
education	13.57 (2.85)				TI
initial vocat. educat.		22.66	24.92	100	TI
voc_educ1		41.46	49.87	76.96	TV
sgen_voc		4.96	6.00	79.76	TV
svoc_voc		9.38	11.79	80.22	TV
tshort_voc		7.81	9.40	75.09	TV
tlong_voc		10.09	10.89	77.93	TV
univ_voc		7.71	9.73	70.89	TV
voc_school		18.66	25.28	64.96	TV
voc_work		8.37	10.81	66.25	TV
voc_altern		7.35	10.78	63.45	TV
voc_forem		13.26	14.84	80.87	TV
voc_train	1.92 (10.85)				TV
exper	19.07 (9.95)				TV
services		53.9	66.65	70.87	TV
professional		32.51	37.00	75.30	TV
BXL		10.96	10.67	95.33	TV
Flanders		49.37	53.78	99.52	TV
Wallon		39.67	36.71	98.10	TV
woman		45.51	46.51	100	TI
nation1		94.99	94.36	100	TI
f_hedu		22.40	22.39	100	TI
f_sedu		42.59	40.80	100	TI
f_ledu		35.01	36.82	100	TI
m_hedu		12.58	12.44	100	TI
m_sedu		45.00	43.47	100	TI
m_ledu		42.42	44.09	100	TI
ms		71.62	73.53	95.01	TV
hlth		85.21	92.91	89.45	TV
age_coh1		14.02	21.23	67.95	TV
age_coh2		56.74	63.65	82.35	TV
age_coh3		29.24	34.21	81.05	TV
Observations	10 811	10 811	10 811	10 811	

From the table above (last column) we see immediately which are the time invariant variables and which exhibit time variation and the magnitude of this variation (column 5). We see for example that a big proportion of individuals (100-76.96= 23.04%) have undertaken vocational education or training during the survey and also many moved to a different profession (24.7% of the sample) or sector of activity (29.13% of the sample).

A considerably high proportion of the individuals in our sample have followed after-school vocational education (41.46%). The data show that generally individuals with higher initial education engage more often in after-school vocational education and training (4.96% of those with general secondary education, 7.81% of those with short-type tertiary education, 10.09% of those with long type tertiary education); a first indication that initial and after school education should be seen as complements. An exception to that are individuals with a university degree (7.71%). We also see that those who followed secondary vocational education at school undertake more often some vocational courses afterwards, compared to individuals with general secondary education.

The majority of individuals prefers courses organised in a school or other academic establishment (18.66%), and second mostly preferred are the courses organised by FOREM and other employment organisations (13.26%). These second ones, target mainly to individuals who have spent some time unemployed and therefore exhibit special features and should be examined differently from the rest three categories. Individuals prefer less vocational education at the workplace (8.37%) or under an alternating scheme (7.35%). These last two vocational course categories are expected to be better related to and connected with the labour market, so at first sight it is somewhat odd that are less chosen. Of course this may have to do with the availability of these courses to interested individuals.

The majority in the sample are married with has good health, Belgian citizens of age 30 to 45, they work in the services sector and live in the Flemish region. Most of them had parents with secondary education, although the percentage of individuals with a low educated mother is very high as well.

From calculations that are not presented here, but are readily available upon request, we conclude that experience and education are negatively correlated, probably because better educated enter the labour market at a later stage in their lives. Finally, age plays an important role in the time variables. More specifically, individuals in the first age cohort spend considerably more time in education and unemployment than the older ones, while individuals in the third cohort are more time inactive and out of the labour market.

6. Estimation results

a. Education and economic performance

We seek here to compare the effect of different education streams (general and vocational, initial and lifelong) on annual and monthly earnings. The expected differences will be attributed either to differences in labour supply or to differences in unemployment time. In that way we will have a whole picture about the effects of education on labour market experiences of individuals. The estimation method followed here is the Hausman- Taylor, which as described above produces efficient and consistent estimators in the presence of correlation between the explanatory variables and the unobserved individual effects. However we present in Appendix 2 the Random Effects and OLS estimators for comparison and further discussion.

After several estimations, with different sets of variables used as exogenous, we arrived at the model HT1 presented in Table 5. As exogenous time variant variables (X_1) we use age cohorts (age_coh1, age_coh3), marital status (ms=1 if married) and health (hlth=1 if health in good state) variables, while exogenous time invariant variables (Z_1) are gender (woman), nationality (nation1=1 if Belgian), father and mother education (f_hedu=1 if father has tertiary education, m_hedu=1 if mother has tertiary education, f_ledu=1 if father has no or primary education and m_ledu=1 if mother has no or primary education). The variables on lifelong vocational education and the interactions between initial and lifelong education ([9]), the variable on experience and its square, the dummies on sector of activity (services=1 if individual works in the services sector) and profession (professional=1 if the individual is a professional) and finnaly the region dummies (Wallon and Brussels) are included in the time-variant endogenous variables (X_2). The main group of interest in the estimations is the time invariant endogenous variables (Z_2) and consists of *education* (years of education) and the *interaction* variables (years of education*1 if the person followed the technical/vocational stream of secondary education).

Education is grouped in initial (general and vocational) and lifelong (completed or undergoing), while we include interactions between initial and lifelong education. All education variables are endogenous, as seen in the previous paragraph with the initial education variables being time invariant. All variables used here are described in detail in Appendix 1. The depended variables are lnw (logarithm of monthly wages), lnwh (logarithm of annual income), lnh* (logarithm of labour supply months) and lnu/h (logarithm of unemployment to employment months ratio).

Table 5: Hausman-Taylor estimations (HT1). (standard errors in parentheses)

Variable	*lnwh*		*lnw*		*lnh**		*lnu/h*	
TI endogenous (Z_2)								
education	0.1852	(0,0420)*	0.0937	(0.0292)*	0.0408	(0.0156)*	-0.0465	(0.0155)*
interaction	-0.0807	(0,0248)*	-0.0575	(0.0153)*	-0.0182	(0.0125)	0.0032	(0.0126)
TV endogenous (X_2)								
voc_educ1	-0.1597	(0.0785)*	-0.0760	(0.0422)**	0.0203	(0.0462)	0.1059	(0.0468)*
sgen_voc	0.2519	(0.0849)*	0.1037	(0.0456)*	0.0182	(0.0499)	-0.1306	(0.0506)*
svoc_voc	0.0897	(0.0818)	0.0646	(0.0439)	-0.0316	(0.0481)	-0.0592	(0.0488)
tshort_voc	0.1608	(0.0815)*	0.0794	(0.0438)**	-0.0382	(0.0479)	-0.1219	(0.0486)*
tlong_voc	0.1757	(0.0810)*	0.1032	(0.0435)*	-0.0193	(0.0476)	-0.0942	(0.0483)**
univ_voc	0.1930	(0.0809)*	0.1117	(0.0435)*	-0.0264	(0.0475)	-0.1104	(0.0482)*
voc_school	0.0176	(0.0147)	0.0032	(0.0079)	0.0108	(0.0086)	-0.0031	(0.0088)

[9] Variables: voc_educ1, sgen_voc, svoc_voc, tshort_voc, tlong_voc, univ_voc, voc_school, voc_work, voc_altern, voc_forem and voc_train

voc_work	0.0163 (0.0211)	0.0055 (0.0113)	0.0076 (0.0124)	-0.0024 (0.0126)
voc_altern	0.0412 (0.0201)*	0.0180 (0.0108)**	0.0077 (0.0118)	-0.0156 (0.0120)
voc_forem	0.0166 (0.0234)	-0.0090 (0.0126)	0.0099 (0.0137)	-0.0154 (0.0139)
voc_train	-0.0010 (0.0003)*	-0.0002 (0.00015)	-0.00017 (0.00017)	0.00064 (0.00017)*
exper	0.0408 (0.0041)*	0.0217 (0.0022)*	0.0142 (0.0023)*	-0.0037 (0.0023)
exper_sq	-0.00092 (0.00009)*	-0.0004 (0.00005)*	-0.00038 (0.00005)*	0.00009 (0.00004)**
services	0.0168 (0.0086)**	0.0077 (0.0046)**	0.0059 (0.0050)	-0.0028 (0.0051)
professional	0.0350 (0.0121)*	0.0049 (0.0065)	0.0227 (0.0070)*	-0.0071 (0.0072)
BXL	0.1309 (0.0923)	0.0456 (0.0499)	0.0709 (0.0532)	-0.0216 (0.0538)
Wallon	-0.0374 (0.0898)	0.0158 (0.0483)	-0.0042 (0.0523)	0.0447 (0.0530)
TI exogenous (Z_1)				
woman	-0.4556 (0.0307)*	-0.3820 (0.0220)*	-0.0443 (0.0102)*	0.0267 (0.0100)*
nation1	-0.0237 (0.0842)	-0.1024 (0.0610)**	0.0174 (0.0256)	-0.0534 (0.0247)*
f_hedu	-0.2238 (0.0554)*	-0.1013 (0.0396)*	-0.0634 (0.0186)*	0.0544 (0.0183)*
f_ledu	0.2600 (0.0576)*	0.1012 (0.0416)*	0.0717 (0.0167)*	-0.0693 (0.0182)*
m_hedu	-0.0677 (0.0500)	-0.0302 (0.0359)	-0.0256 (0.0164)	0.0115 (0.0161)
m_ledu	0.1913 (0.0529)*	0.0762 (0.0383)*	0.0723 (0.0167)*	-0.0367 (0.0163)*
TV exogenous (X_1)				
ms	0.0281 (0.0170)**	-0.0133 (0.0097)	0.0093 (0.0078)	-0.0250 (0.0077)*
hlth	0.0074 (0.0110)	0.0105 (0.0060)**	-0.00068 (0.0065)	0.0012 (0.0066)
age_coh1	0.0036 (0.0179)	0.0024 (0.0097)	0.0285 (0.0104)*	0.0249 (0.0105)*
age_coh3	0.0369 (0.0163)*	0.0187 (0.0088)*	0.0159 (0.0095)**	-0.0049 (0.0096)
constant	10.601 (0.6273)*	9.6505 (0.4265)*	1.7833 (0.2521)*	0.7512 (0.2510)*
obs. number	10 811	10 811	10 811	10 811
Hausman test statistic	$\chi^2(21)$=6.09 (0.9994)[10]	$\chi^2(21)$=1.21 (1.000)	$\chi^2(21)$=3.30 (1.000)	$\chi^2(21)$= 4.49 (0.9999)

(*) Significant at 95% level.
(**) Significant at 90% level

The relationship between the coefficients in all four models we estimated here is given by equation (2.4) in part 2. It is easy to check that for all estimated coefficients equation (2.4) approximately holds. Concerning more specifically education coefficients we see that the difference (0.1852-0.0937-0.0408+(-0.0465))= –0,0042, which is very small and not statistically significant. These results confirm the initial hypothesis that the return of education on earnings is composed by the return on monthly wages and employment time or stated differently, that the difference between annual and monthly earnings education coefficients is attributed to time in the labour market and unemployment.

We have also tested for correlation between the education (time-invariant) variables and the unobserved individual effect using the Hausman specification test. Each one of the four models above is tested against a fixed effects model. The Hausman statistics are given in the last row of table 5. As we can see the null hypothesis of no correlation between the explanatory variables and the unobserved individual effects is not rejected for any of the four models. This implies that we have put each variable in the correct group (X_1, X_2, Z_1, and Z_2) and that our instruments (X_1 group) are legitimate.

In a broader model that would also incorporate those who did not receive any income from work because they are either unemployed or out of the labour market throughout the year we expect both education coefficients and the gap between monthly and annual income to be different than the ones calculated here. However whether these expected differences will be negative or positive is something that depends on the quantity and quality of those in unemployment, the type and level of unemployment benefits etc. If for example unemployed are mostly those with low qualifications we could expect larger education coefficients; however if the replacement ratio (b/w) is larger for less educated individuals, then education coefficients could decrease.

[10] p-values in parentheses

Compared to initial vocational education, general education implies better incomes, less unemployment, and more time in the labour market, although the differences are not statistically significant for the time equations. Remember that here we have already dealt with the problem of correlation between schooling and individual abilities (the econometric model we are using does just that), so the reasoning that vocational education at school is chosen and followed by less advantaged pupils cannot apply. The labour market seems to value more general education skills and knowledge (at a secondary education level) than specific technical or vocational skills. These results are somewhat different compared to the results on US surveys. There the evidence suggest that individuals who do not continue to higher education are better off when they receive vocational rather than general education courses while at high school (e.g. Mane, 1999). However these findings are specific to US education system, which in general gives fewer opportunities for post-school non-university education than Eurorean education systems. In surveys based on European data however, it shows that individuals who have followed general education courses while at secondary education do better than those who graduate from the vocational streams of secondary education (Cooke, 2003). UK surveys (Dearden et al, 2000, Conlon, 2001) also find differentials in earnings premia for general and vocational education in favour for general education.

The effects of vocational after-school education vary considerably depending on the initial education a person has received. Those with very low initial education levels (primary education or less) seem even to be harmed when they undertake a vocational course. This somewhat odd result could imply that individuals with very low initial education are not in a position to understand and successfully process the knowledge and skills offered by a vocational course. It can also imply that for these individuals their presence in the labour market is of major importance, so when they leave in order to follow vocational training they loose. On the other hand there are considerable gains from after school education for those who have completed general secondary education and university education. In the case of university graduates who already have a high level of academic knowledge and have developed analytical skills, vocational education provides them with practical knowledge and skills needed in the labour market. The gains, both in incomes and in unemployment time are smaller for those with non-university tertiary education, but still significant. Finally, vocational after-school education does not really affect individuals who already completed vocational education at school. Similar results are found in Cooke (2003) where post-school vocational education is much more profitable when individuals have completed the "Abitur" rather than vocational secondary education; however for Britain there is evidence that vocational qualifications while at school receive their highest return when they are followed by vocational qualifications after school, but lose some of their value if followed by general (academic) education (Dearden et. al, 2000).

A first conclusion that we can draw here is that at least up to secondary education level, general skills and knowledge are more useful to students, while specialisation and training is something that should come at a later stage in life. A second conclusion is that, because continuing vocational education is usually very market-oriented and provide with applied and specific skills, it is of more use to those who have high levels of theoretical knowledge and have never had received similar vocational skills before (such as university graduates).

Concerning the type of vocational education that is more valued in the labour market, courses that combine both theoretical knowledge and practice at the workplace are the only ones that have a significant positive impact on annual and monthly incomes. The rest of the courses have positive effects on incomes but these are not different from 0. Furthermore the type of vocational education does not affect significantly unemployment time, while labour supply is only affected by years of initial education and not by after school education, types of courses etc.

The rest of the explanatory variables do not seem to exhibit any strange features. Working as a professional and at the services sector implies better income and less unemployment compared to other occupations and sectors. On the other hand, labour market prospects seem to get better with age, as we see that unemployment hurts especially those in the first age cohort (18-30). Contrary to what someone might think, those with Belgian nationality seem to earn less than non-Belgians. This could be explained if one takes into account the large number of well paid administrators from the EU and other countries who work in the various EU and international organisations (both private and public) that exist in Belgium.

In results that we do not present here but are available upon request, we have estimated economic returns of education separately for males and females. There are a few points to comment. First the effects of initial general education are in general higher for women than for men for all of our four models. However the gender difference is very small in the wage model while it is substantially larger in the labour supply and unemployment models. This implies that education is a much more important factor for participation in the labour market for women than it is for men but it does not affect very differently the wages of men and women. Secondly, vocational initial education implies lower wages and consequently earnings for both genders, but it hurts women more than men. Lifelong vocational education does not have affect very different men and women, even though the effects for women are somewhat larger. A final point that is not apparent in the both genders' regressions is that vocational courses organised by Forem significantly reduce unemployment for men but not for women. The rest of the results and the analysis do not change.

b. Hausman Taylor, Random effects and OLS estimators

The use of Hausman Taylor estimation method has produced consistent estimators and corrected the problems of endogeneity of our main explanatory variables, namely the education variables. This is shown by the Hausman test statistics, where we do not reject the null hypothesis of no correlation for any of the four models (lnwh, lnw, lnh^* and lnu/h). In Table 6 we report together the years of initial *education* and *interaction* coefficients (Z_2 group) estimated with HT, RE and OLS. This will facilitate the comparison between the three and further discussion. The RE and OLS coefficients for the rest of the explanatory variables are reported in Appendix 2, Tables 2.A and 2.B.

Table 6 HT, RE and OLS education coefficients (standard errors in parentheses; the rest of the explanatory variables are omitted)

Variable	*lnwh*	*lnw*	*lnh^**	*lnu/h*
	Education coefficient estimated by Hausman Taylor (HT)			
education	0.1852 (0,0420)*	0.0937 (0.0292)*	0.0408 (0.0156)*	-0.0465 (0.0155)
interaction	-0.0807 (0,0248)*	-0.0575 (0.0153)*	-0.0182 (0.0125)	0.0032 (0.0126)*
Hausman test statistic	$\chi^2(21)$=6.09 (0.9994)[11]	$\chi^2(21)$=1.21 (1.000)	$\chi^2(21)$=3.30 (1.000)	$\chi^2(21)$= 4.49 (0.9999)
	Education coefficient estimated by Random effects (RE)			
education	0.0546 (0.0037)*	0.0487 (0.0026)*	0.0016 (0.0015)	-0.0039 (0.0018)
interaction	-0.0071 (0.0022)*	-0.0075 (0.0015)*	-0.00006 (0.0009)	-0.00031 (0.00096)*
Hausman test statistic	$\chi^2(20)$=125.45 (0.000)	$\chi^2(20)$=196.63 (0.000)	$\chi^2(20)$=32.91(0.0345)	$\chi^2(20)$= 30,27 (0.0656)
	Education coefficient estimated by Ordinary Least squares (OLS)			
education	0.0489 (0.0021)*	0.0454 (0.0016)*	0.0004 (0.0009)	-0.0031 (0.0009)*

[11] p-values in parentheses

interaction	-0.0079 (0.0012)*	-0.0076 (0.0009)*	-0.00036 (0.0005)	0.00006 (0.00055)
Observations	10 811	10 811	10 811	10 811

(*) Significant at 95% level.
(**) Significant at 90% level

As is shown in Table 6, the Hausman test does not reject the hypothesis of no correlation for all four models that have been estimated by the HT method. On the contrary the Hausman statistics for our RE models are very much higher and the null hypothesis is rejected for all of them. The proximity between RE and OLS coefficients implies that OLS estimators are not consistent here either.

Before going any further in comparing RE (and OLS) to HT we would like to discuss briefly a relatively recent block of literature that questions either the Mincerian framework or the usual econometric techniques (OLS, IV) as the most appropriate ways to estimate the returns to education. First of all there is the critique that the schooling coefficient in the Mincer equation equals the marginal internal rate of return to education only under certain (quite restrictive) assumptions. More specifically it is assumed that education and experience are separable, foregone earnings are the only costs of education, the earnings measure used captures the whole benefit of the investment of education etc. (Björklund and Kjellström, 2000). Heckman, Lochner and Todd (2003) argue that the data from the 1940's and 1950's provide some support for the Mincer model but data from later decades are inconsistent with it. When they relax functional form assumptions of the Mincer equation and also account for nonlinearity in schooling, taxes and tuition and nonseparability between education and experience they conclude that the internal rate of return of education should be much higher than predicted by the Mincer model.

On the other hand, on the issue of endogeneity of education, Heckman and Vytlacil (2000) argue that education and ability are literally inseparable. This implies that it is impossible to separate and estimate the main effects of education, even if ability were fully observable. Moreover in the literature, people mostly try to deal with the correlation between cognitive skills and education, and usually oversee the important role of non-cognitive skills and how they affect educational and labour market performance (Carneiro and Heckman, 2003). The non-separability between education and abilities and the possible non-linearity of the wage equation to education make not only OLS but also IV techniques ill-equiped to estimate accurately the returns to education. More sophisticated techniques such as non-parametric or semi-parametric econometric models or structural dynamic models of schooling decisions along with wage regressions that account for skill-heterogeneity, non-linearity in schooling, education-experience separability and heteroskedasticity are more appropriate (Belzil and Hansen, 2002 and Belzil, 2004). These techniques usually result to education estimates, which are much smaller than OLS/ IV estimates (Belzil and Hansen, 2002 report average returns to education from less than 0.010 to 0.12 and Belzil, 2004 reports similar lower and upper bounds).

As appealing as those critiques to the Mincer model and the OLS/ IV estimation methods may be, the development of new models and estimation techniques is still at a very early stage, the empirical work is still limited and the validity of results remains to be confirmed. We can not overlook the fact that depending on the issue each survey addresses, the results are so different; for example Heckman, Lochner and Todd (2003) who work on a more flexible functional form of the Mincer equation, report much higher education coefficients than OLS/ IV estimates. Belzil and Hansen (2002) and Belzil (2004) on the other hand arrive at substantially lower estimates of the return to education. So further research needs to be done before dismissing the techniques and models used up to date. On the other hand, the relative simplicity of the Mincer equation as well as of the OLS/ IV techniques makes them even more appealing.

Going back to our discussion on the relationship between RE (OLS) and HT estimators we see that the HT estimators here are considerably higher than the Random effects and OLS estimators implying that when we are using estimation methods that control for individual unobserved effects, the true effect of education on earnings is much larger. This is consistent with the literature on HT estimation (HT (1981), Cornwell and Rupert (1988), Baltagi and Khanti-Akom (1990) etc), but also with the literature on traditional IV estimation (see Card, 1999 for a review). The higher estimates reported in the literature when estimations methods to control for unobserved individual effects are used is somewhat counter-intuitive. RE (and OLS) estimators should be upward biased if we assume that there is a positive correlation between individual characteristics and amount of schooling, but the empirical results suggest otherwise.

Card (1999) summarises a number of potential explanations for the gap between OLS and the traditional IV estimators (e.g. measurement error that biases downwards OLS estimators, "publication bias" i.e. because researchers search for a statistically significant IV estimator, they are more likely to select a model specification that produces a large point estimate of the return of education). Another explanation that becomes popular in the literature is the so-called LATE (Local Average Treatment Effect) interpretation of the IV; according to that, each instrument used (proximity to university, family background, compulsory schooling laws etc) affects a different group of people and in each case affects those who choose more (or less) schooling because of the instrument; these are the *compliers*. However the LATE interpretation applies only in the case of a unique instrument. When using one or more instruments (as here) for education we can not define a unique group that is affected.

So, how can the gap between Random Effects/ OLS and HT estimators be explained? In the literature, while all realise the existence of this gap, which is even larger than the OLS/ traditional IV gap, no explanation seems to be provided. In their seminal article Hausman and Taylor refer to the issue with the following: "All methods which control for correlation with the latent individual effects increase the schooling coefficient over those which do not; and this is certainly not the direction that many people concerned about ability bias would have expected" (HT, 1981 pp. 1393). The references in the rest of the literature do not seem more convincing.

An economic intuition (but not econometric) however for the gap between HT and RE/ OLS would be that the true effects of education are underestimated in case that education acts as a cover and substitute for low ability, correcting individual deficiencies. This of course would imply that the majority of individuals have less than average levels of ability or that individuals with exceptional IQ levels, ambition or talents leave education earlier in order to pursue their goals in the labour market or that they follow forms of education that are not registered or followed by surveys (e.g. out of school music lessons, or sport practices). Any of the above could seem plausible, but more certain answers require further research.

A second issue to be discussed is that the gap between annual and monthly earnings and consequently the effects of education on labour supply and unemployment are much larger than implied both by RE / OLS estimations and descriptive statistics. Part of this increase can of course be attributed to the general "inflation" of HT estimators, but we see that controlling for ability does not affect education coefficients for all variables the same. Assuming that we accept that the HT estimators are consistent compared to RE (and OLS) estimators and more efficient than the within estimators, the true gap between annual and monthly earnings is much larger. In other words, the true effect of education on earnings is much larger and broader than when we consider only the amount of money a person receives per unit of time; the time the person receives it plays an important role that should not be overseen.

7. Conclusions

There are not many studies that estimate the effect of education on earnings and employment time in the way that is being done here. We have shown that education improves significantly not only the income of individuals but also the experiences they have in the labour market. This improvement as estimated here is much larger than traditional estimation methods predict. A plausible explanation is that education is a substitute for low ability and that highly talented individuals enter the labour market earlier; however these reasons should be further explored.

The results also shed some light in the debate between general or vocational education. The gains, both in terms of income and employment, are higher for those who have followed general education at school, even after controlling for individual ability. It is therefore important for individuals to acquire and develop general skills while at school and while they are at the stage in life when they form their personality. Vocational education and training is a better choice after school and even after the beginning of one's work life.

The assumption of complementarities between initial and after-school education is confirmed in the paper. The positive effects of after school education are significantly higher for those with higher initial education and are also larger for those with general initial education.

Based on the results presented here, individuals should be encouraged to receive education, because it implies better income prospects and more successful labour market experiences. When at school, children are better to acquire general knowledge and develop their personal skills and competences. Specialisation and practical skills further improve individual prospects but are more useful to individuals who have learned how to process knowledge and information and have strong theoretical background. In a word, we could say that educational policies should take a turn towards forming complete personalities of children at school and building solid backgrounds for them and then providing them with continuous and updated specialised and labour-market tailored knowledge.

References

Anemiya, T., MaCurdy, T,E., "Instrumental Variable estimation of an errors components model", Econometrica, vol 54(4), p. 869-880

Arellano, M., (2003) "Panel Data Econometrics", Oxford University Press

Ashenfelter, O., Ham, J., (1979) "Education, unemployment and Earnings", Journal of Political Economy vol. 87(5), part 2 October 1979

Baltagi, B.H, Bresson, G., Pirotte, A., (2003), "Fixed effects, random effects or Hausman-Taylor? A pretest estimator", Economics letters, vol. 79, p. 361-369

Baltagi, B.H., Khanti-Akom, S., (1990), "On efficient estimation with panel data: An empirical comparison of instrumental variables estimators", Journal of Applied econometrics, Vol 5(4), p. 401-406

Bartel, A.P., Lichtenberg, F. R., (1987) "The comparative advantage of educated workers in implementing new technology", The Review of Economics and Statistics, Vol. 69 (1), February 1987

Becker, G.S. (1964) "Human capital: A theoretical and empirical analysis with special reference to education", Chicago, University of Chicago Press.

Belzil, C., "On the specification of Mincerian Wage regressions with Heterogeneity, Non-Linearity, Non-Separability and Heteroskedasticity", IZA Discussion Paper No. 1083, March 2004

Belzil, C., Hansen, J., (2002) "Unobserved ability and the return to schooling", IZA Discussion Paper, No. 508, May 2002

Benson, C.S., (1997) "New-vocationalism in the United States: potential Problems and outlook", Economics of Education Review, Vol. 16(3) p. 201-212

Björklund, A., Kjellström, C., (2000) "Estimating the return to investments in education: how useful is the standard Mincer equation?" Economics of Education Review, vol. 21 p. 195-210

Breusch, T.S, Mizon, G.E., Schmidt, P., (1989), "Efficient estimation using Panel data", Econometrica, Vol 57(3), p.695-700

Brunello, G. (2001) "On the complementarity between education and training in Europe", IZA Discussion Paper, No.309, June 2001

Card, D., (1999), "The causal effect of education on earnings", Handbook of labor Economics, Volume 3, Ashenfelter and Card (eds.), vol 3.

Carneiro, P., Heckman, J., (2003) "Human Capital Policy", National Bureau of Economic Research, Working Paper No. 9495, February 2003

Carnoy, M., (1997) "The Great Work Dilemma: Education, Employment and Wages in the global Economy", Economics of Education Review, Vol. 16(3) p. 247-254

CEDEFOP (2001), "Vocational education and training in Belgium- Brief description", Cedefop, Panorama series, Luxembourg 2001;

Chowdhury, G., Nickell, S., (1985) "Hourly earnings in the United states: Another look at unionization, Schooling, sickness and Unemployment using PSID data", Journal of Labor Economics, Vol. 3(1) Part 1, p. 38-69.

Conlon, G., (2001), “The differential in Earnings Premia between Accademically and Vocationally trained males in the united Kingdom”, Centre for Economic Performance, Discussion Paper CEEDP0011, June 2001.

Cooke, L.P. (2003) “A comparison of initial and early life course earnings of the German secondary education and training system” Economics of education review, vol 22, p.79-88

Cornwell, C., Rupert, P., (1988), “Efficient estimation with Panel data: An empirical comparison of Instrumental variables estimators”, Journal of Applied Econometrics, Vol 3(2), p. 149-155

Dearden, L., McIntosh, S., Myck, M., Vignoles, A., (2000), “The returns to academic and Vocational Qualifications in Britain”, Centre for Economic Performance, discussion paper no. CEEDP0004, November 2000.

Eurydice (2001) “The education system in Belgium, (2000-2001)”, Eurybase- the information database on education systems in Europe www.eurydice.orgEurybase/frameset_eurybase.html

Gardner, R., (1998), “Unobservable individual effects in unbalanced panel data”, Economics letters, vol. 58, p. 39-42

Grubb, W.N. (1997) “The Returns to Education in the Sub-baccalaureate Labor Market”, 1984-1990”, Economis of education Review, vol 16(3) p. 231-245

Hausman, J.A., (1978), “Specification tests in econometrics”, Econometrica, vol. 46(6), p. 1251-1271.

Hausman, J.A., Taylor, W.E., (1981), “Panel Data and unobservable individual effects”, Econometrica, Vol 49 (6), p. 1377-1398

Heckman, J., Vytlacil, E., “Identifying the role of cognitive ability in explaining the level of and change in the return to schooling”, National Bureau of Economic Research, Working Paper No. 7820, August 2000

Heckman, J., Lochner, L., Todd, P., (2003) “Fifty years of Mincer earnings regressions” National Bureau of Economic Research, Working Paper No. 9732, May 2003

Holzer, H. J. (1996): “What employers want: Job prospects for less educated workers”, New York: Russell Sage Foundation, 1996

Im, K.S, Ahn, S.C., Schmidt, P., Wooldridge, J.M., (1999) “Efficient estimation of panel data models with strictly exogenous explanatory variables”, Journal of Econometrics, vol. 93, p. 177-201

Mane, F., (1999) “Trends in the payoff to academic and occupation-specific skills: the short and medium run returns to academic and vocational high-school courses for non-college-bound students”, Economics of Education Review, Vol. 18 p. 417-437

McIntosh, S., (2004) “Further Analysis of the Returns to Academic and Vocatioanl qualifications” Centre for Economic Perormance Discussion paper CEEDP0035, January 2004.

Mincer, J., (1991) “Education and unemployment”, NBER working paper 3838

Mincer, J., (1974) “Schooling, experience and earnings”, Columbia University Press, New York

Nickell, S., Bell, B., (1995) “The collapse in demand for the unskilled and unemployment across the OECD”, Oxford review of Economic Policy, vol. 11(1), spring 1995

OECD, (2003), "Education at a glance", Paris 2003

Pereira, T.P, Martins, S.P (2001) "Returns to Education and Wage Equations", IZA Discussion Paper No. 298, June 2001

Stern, D., Finkelstein, N, Urquiola, M., Cagampang, H., (1997) "What Difference does it make if school and work are connected? Evidence on Co-operative Education in the United states" Economics of Education Review, Vol. 16(3) p. 213-229.

Wooldridge, J., (2001) "Econometric Analysis of cross section and panel data", MIT press

Appendix 1
Construction of variables and connection to the PSBH questionnaire.

1.A Education variables

Along with the detailed description of the construction of our variables we are also mentioning the relevant PSBH questions we used. The codes of PSBH questions are of the form **p**ixxxx and **w**i**a**xxxx (where i= 4…10 the wave number and xxxx is a four digit number).

1.A.a Initial education (general and vocational)

In the PSBH, individuals are asked which is the highest education level they have completed, (questions p0i3820), at what age they have completed their full time education (p0i3815), and at what age they have completed the highest education level (p0i3840). We decided to use the information on the highest completed level of education instead of the age of education completion. This decision was based mainly in two reasons: firstly the self -reported age of education completion is highly inaccurate. Individuals do not report any breaks they had from school to the labour market and then back to school, while many individuals report only the highest education level completed and not their age at the time. Secondly, we want to avoid the problems we would have with repeated education years or years that did not led to a degree. If these years were included at the education variable then the true effect of education would be seriously underestimated since less able individuals (those who usually repeat school years) would appear to have more education and yet small incomes. Moreover, with the econometric model used here we control for individual ability.

We also tried to incorporate in the education variable two changes in the Belgian education system, which are described in detail in part 2.a of this paper. The first change affected secondary education. The new system begun implementing in 1971, even though there has been a long transition period when both old and new systems were valid. However here we assume that the new system was fully implemented in 1971, so it affected individuals born after 1959 (=1971-12). Students who entered secondary education up to 1971 have followed two cycles, of a three –year duration each. These cycles were called lower and upper secondary education respectively. After the reform of 1971, secondary education comprises of three (two- years each) cycles, so now those who have completed lower secondary education have had 4 years of secondary education compared to 3 before 1971).

The second reform affected tertiary short type education. Up to 1990 (individuals who were born till 1972) short type tertiary education was only two- year courses. After 1990 it lasts 3 years (and 4 in some specialisations).

Taking all these into consideration "education" variable then takes the following values:

i) education=6 if individual has finished (or not) primary education (pi3820==1)
ii) education=9 if individual was born before 1959 (so went to secondary education before the reformed system came into force in 1971) and has finished lower secondary education, either general, artistic, technical or professional (pi3820=2, 3, 4, 5).
iii) education=10 if individual was born after 1959 (so went to secondary education after the reformed system came into force in 1971) and has finished lower

secondary education (cycles 1 and 2 of the reformed system), either general, artistic, technical or professional (pi3820=2, 3, 4, 5).

iv) education=12 if individual has finished upper secondary education, either general, artistic, technical or professional (pi3820=6, 7, 8,9), regardless of his year of birth.

v) education=14 if individual was born before 1972 (so followed tertiary education before 1990) and has completed a short type tertiary education course, either pedagogical, economic or other (pi3830=1, 2, 3).

vi) education=15 if individual was born after 1972 (so followed tertiary education after 1990) and has completed a short type tertiary education course, either pedagogical, economic or other (pi3830=1, 2, 3).

vii) education=16 if individual has followed a long-type non-university tertiary education course (pi3830=4).

viii) education=14 if individual has acquired a “diplôme de candidatures” (pi3830=5).

ix) education= 16 if individual has acquired a “diplôme de licences” (pi3830=6).

x) education= 18 if individual has acquired a “doctorat, maîtrise ou diplôme de 3ème cycle” (pi3830=7).

Moreover, to distinguish between general and vocational secondary education we created a dummy variables: *voc_degree=1* if individual has completed technical or vocational secondary education. The variable “*interaction”* that is used in the regressions is the product of *“education”* and *“voc_degree”* variables and measures the effect of years in vocational secondary education.

1.A.b Vocational (after school) education

- *“voc_educ1”*=1 if pi3850=1. Indicates whether the individual ever had at least one vocational course.
- *voc_school=1* if pi3870=1. The vocational course was organised in a school or similar education establishment (college etc)
- *voc_educ_work=1* if pi3890=1. The individual has followed a vocational education at the workplace.
- *voc_educ_altern=1* if pi3880=1. The individual has completed a vocational course under an alternationg system (school and workplace).
- *voc_educ_forem=1* if pi3910=1. The individual has followed some course within an employment programme (e.g. FOREM, ORBEM etc).

It is obvious that an individual may have followed more than one types of vocational courses and all these variables are time-variant by construction of the sample.

To capture the effect of after school education by level of education we also created the following variables, that measure interactions:

- *sgen_voc=1* if voc_educ1=1 & the individual has completed secondary general education
- *svoc_voc=1* if voc_educ1=1 & the individual has completed secondary vocational education
- *tshort_voc=1* if voc_educ1=1 & the individual has completed tertiary short type education
- *tlong_voc=1* if voc_educ1=1 & the individual has completed tertiary long type education
- *univ_voc=1* if voc_educ1=1 & the individual has completed university education.

Finally the variable *"voc_train"* measures the duration in weeks of vocational course undertaken during the period of the survey. We constructed it combining information from pi1540 ("have you followed a vocational education course since last year?"), pi1550("what was the total duration of that course), pi1590 ("starting year of the course"), pi1610 ("finishing year of the course"), pi1600 ("starting month of the course") and pi1610 ("finishing month of the course").

1.B Income variables

We used income information for the previous economic year. PSBH has questions on both gross and net monthly amounts. However, following the common practice in the estimation of private returns of education we are going to use only net income, since gross income includes also social returns.

The monthly wage variable (only for those who received income from work and not from business activity or investment) consequently is:
net monthly income= pi1890
In case pi1890 is missing and only gross monthly wage (variable pi1880) is available, we calculated net monthly wage as a second degree function of gross monthly income of the form: *net monthly wage=a(gross monthly wage)+b(gross monthly wage)*^{2}*+c*.
By regressing net wage on gross wage and its square the estimates for a and b were: $\hat{\alpha}$=0,47, $\hat{b}$=6,86e-08 and $\hat{c}$=12.354.

The annual income variable is:
net monthly income × pi1900, where pi1900= number of months that a person received reported net monthly income.
Finally, in order to make income comparable from year to year we expressed all income variables in fixed 1995 prices using relevant data from "Belgostat" (www.belgostat.be).

1.C Time variables

In PSBH, individuals are asked about their main activity month by month (questions pi1710, pi1720, pi1730….pi1820). By summing up their answers I calculated the months an individual has spent in work, in unemployment, in education and in inactivity each year. Moreover, in the estimations we are using only individuals who report months in employment equal to pi1900 (months received reported net monthly income). In such way we want to minimise errors in the calculation of annual income and employment/ unemployment months.

1.D Other variables

- *Exper*= Age-pi3770, where pi3770 is the age of the first regular full-time job. Experience is measured in years.
- *services*=1 if individual is occupied in the service sector. The classification into sector of activity was done using the NACE classification system.
- *professional=1* if the individual belongs to "legislators, senior officials and corporate managers" or "professionals". The classification followed the ISCO classification system.
- *woman=1* if wia006=2 ("Sex")
- *nation=1* if pi3550=1 ("What is your present nationality?")

- *father education*= wi060
- *mother education*= wi061
- *ms=1* if pi3620=1 ("What is your present official family situation?")
- *hlth=1* if pi3390=1 or 2 ("How is your general health situation?").

Appendix 2

Table 2.A Random effects estimators (standard errors in parentheses)

Variable	*lnwh*	*lnw*	*lnh**	*lnu/h*
education	0.0546 (0.0037)*	0.0487 (0.0026)*	0.0016 (0.0015)	-0.0039 (0.0018)*
interaction	-0.0071 (0.0022)*	-0.0075 (0.0015)*	-0.00006 (0.0009)	-0.00031 (0.00096)
voc_educ1	-0.0665 (0.0528)*	-0.0248 (0.0336)	0.0115 (0.0251)	0.0603 (0.0257)*
sgen_voc	0.1254 (0.0563)*	0.0501 (0.0360)	-0.0015 (0.0266)	-0.0726 (0.0272)*
svoc_voc	0.0317 (0.0557)	0.0331 (0.0354)	-0.0239 (0.0266)	-0.0259 (0.0271)
tshort_voc	0.0809 (0.0551)	0.0311 (0.0350)	-0.0146 (0.0262)	-0.0725 (0.0267)*
tlong_voc	0.0306 (0.0553)	0.0154 (0.0351)	-0.0121 (0.0264)	-0.0474 (0.0270)**
univ_voc	0.1256 (0.0555)	0.0778 (0.0352)	-0.0121 (0.0266)	-0.0609 (0.0272)*
voc_school	0.0065 (0.0128)	-0.0033 (0.0076)	0.0041 (0.0066)	-0.0071 (0.0069)
voc_work	0.0227 (0.0179)	0.0107 (0.0107)	0.0019 (0.0091)	-0.0059 (0.0094)
voc_altern	0.0206 (0.0172)	0.0098 (0.0103)	0.0042 (0.0088)	-0.0043 (0.0091)
voc_forem	-0.0335 (0.0180)**	-0.0275 (0.0111)*	-0.0048 (0.0089)	0.0047 (0.0091)
voc_train	-0.0010 (0.0003)*	-0.0002 (0.00015)	-0.00015 (0.00014)	0.0007 (0.0001)*
exper	0.0435 (0.0025)*	0.0241 (0.0016)*	0.0155 (0.0012)*	-0.0030 (0.0012)*
exper_sq	-0.0008 (0.00006)*	-0.00039 (0.000037)*	-0.00036 (0.00003)*	0.00005 (0.000028)**
services	0.0104 (0.0069)	0.00048 (0.0041)	0.0063 (0.0036)**	-0.0078 (0.0037)*
professional	0.0774 (0.0104)*	0.0322 (0.0062)*	0.0228 (0.0053)*	-0.0111 (0.0055)*
BXL	0.0238 (0.0268)	-0.00014 (0.0184)	0.0208 (0.0118)**	0.0069 (0.012)
Wallon	-0.0603 (0.0179)*	-0.0368 (0.0127)*	-0.0041 (0.0077)	0.0152 (0.0077)*
woman	-0.4039 (0.0171)*	-0.3503 (0.0122)*	-0.0338 (0.0072)*	0.0197 (0.0073)*
nation1	-0.0435 (0.0376)	-0.0537 (0.0269)*	-0.0143 (0.0159)	-0.0275 (0.0161)**
f_hedu	-0.0158 (0.0238)	0.0087 (0.0172)	-0.0092 (0.0101)	0.0132 (0.0101)
f_ledu	0.0100 (0.0233)	-0.0121 (0.0167)	0.0039 (0.0098)	-0.0166 (0.0099)**
m_hedu	0.0404 (0.0286)	0.0267 (0.0206)	0.0048 (0.0121)	-0.0087 (0.0122)
m_ledu	-0.0545 (0.0226)*	-0.0267 (0.0163)**	0.0035 (0.0095)	0.0295 (0.0096)*
ms	0.0362 (0.0136)*	-0.0070 (0.0085)	0.0108 (0.0064)**	-0.0285 (0.0066)*
hlth	0.0182 (0.0100)**	0.0140 (0.0059)*	0.0021 (0.0052)	-0.0039 (0.0055)
age_coh1	-0.0145 (0.0158)	-0.0013 (0.0094)	0.0165 (0.0081)*	0.0325 (0.0084)*
age_coh3	0.0331 (0.0145)*	0.0244 (0.0086)*	0.0068 (0.0075)	0.0052 (0.0077)
constant	12.210 (0.0737)*	10.070 (0.0516)*	2.30 (0.0323)*	0.1468 (0.0327)*
obs. number	10 811	10 811	10 811	10 811
Hausman test statistic	$\chi^2(20)$=125.45 (0.000)[12]	$\chi^2(20)$=196.63 (0.000)	$\chi^2(20)$=32.91(0.0345)	$\chi^2(20)$= 30,27 (0.0656)

(*) Significant at 95% level.
(**) Significant at 90% level

Table 2.B OLS estimators (standard errors in parentheses)

Variable	*lnwh*	*lnw*	*lnh**	*lnu/h*
education	0.0489 (0.0021)*	0.0454 (0.0016)*	0.0004 (0.0009)	-0.0031 (0.0009)*
interaction	-0.0079 (0.0012)*	-0.0076 (0.0009)*	-0.00036 (0.0005)	0.00006 (0.00055)
voc_educ1	0.0195 (0.0381)	0.0426 (0.0278)	-0.0049 (0.0157)	0.0181 (0.0176)
sgen_voc	0.0256 (0.0384)	0.0038 (0.0280)	-0.0010 (0.0158)	-0.0228 (0.0171)
svoc_voc	0.0389 (0.0399)	0.0372 (0.0291)	-0.0013 (0.0164)	-0.0030 (0.0178)
tshort_voc	0.0258 (0.0386)	-0.0123 (0.0282)	0.0136 (0.0159)	-0.0245 (0.0172)
tlong_voc	-0.0931 (0.0395)*	-0.1179 (0.0288)*	0.0101 (0.0163)	-0.0147 (0.0176)
univ_voc	0.0222 (0.0406)*	0.0602 (0.0295)*	0.0144 (0.0167)	-0.0175 (0.0181)
voc_school	-0.0165 (0.0134)	-0.0234 (0.0098)*	-0.0036 (0.0055)	-0.0105 (0.0059)**
voc_work	0.0493 (0.0164)*	0.0481 (0.0119)*	-0.00071 (0.0067)	-0.0019 (0.0073)
voc_altern	-0.0136 (0.0167)	-0.0059 (0.0121)	-0.0023 (0.0068)	0.054 (0.0074)
voc_forem	-0.806 (0.0145)*	-0.0533 (0.0106)*	-0.0110 (0.0059)**	0.0163 (0.0064)*
voc_train	-0.0013 (0.00037)*	-0.0003 (0.00027)	-0.00017 (0.00015)	0.00079 (0.00016)*
exper	0.0359 (0.0018)*	0.0227 (0.0013)*	0.0106 (0.00076)*	-0.0026 (0.00082)*
exper_sq	-0.0006 (0.00004)*	-0.0003 (0.00003)*	-0.00024 (0.000017)*	0.000035(0.000019)**
services	0.0110 (0.0083)	-0.158 (0.0061)*	0.0097 (0.0034)*	-0.0172 (0.0037)*
professional	0.1556 (0.0097)*	0.1339 (0.0071)*	0.0115 (0.0039)*	-0.0102 (0.0043)*
BXL	0.0218 (0.0141)	0.0219 (0.0103)*	0.0097 (0.0058)**	0.0098 (0.0063)
Wallon	-0.0179 (0.0089)*	-0.0157 (0.0065)*	0.0012 (0.0036)	0.0035 (0.0039)

[12] p-values

woman	-0.3508	(0.0083)*	-0.3139	(0.0061)*	-0.0227	(0.0034)*	0.141	(0.0037)*
nation1	-0.0573	(0.0187)*	-0.0761	(0.0137)*	-0.0060	(0.0077)	-0.0249	(0.0084)*
f_hedu	-0.0237	(0.0112)*	-0.0032	(0.0082)	-0.0065	(0.0046)	0.0139	(0.0050)*
f_ledu	-0.0147	(0.0110)	-0.0219	(0.0080)*	0.0029	(0.0045)	-0.0044	(0.0049)
m_hedu	0.0304	(0.0135)*	0.0205	(0.0098)*	0.0063	(0.0055)	-0.0036	(0.0060)
m_ledu	-0.0339	(0.0107)*	-0.0197	(0.0098)*	0.0021	(0.0044)	0.0162	(0.0048)*
ms	0.0033	(0.0095)	-0.0172	(0.0067)*	-0.0002	(0.0038)	-0.0207	(0.0041)*
hlth	0.0616	(0.0114)*	0.0382	(0.0083)*	0.0116	(0.0047)*	-0.0118	(0.0050)*
age_coh1	-0.0525	(0.0166)*	-0.0176	(0.0121)	-0.0023	(0.0068)	0.0326	(0.0073)*
age_coh3	0.0431	(0.0147)*	0.0535	(0.0107)*	-0.00076	(0.0060)	0.0096	(0.0065)
constant	12.3165	(0.0431)*	10.1259	(0.0296)*	2.3596	(0.0178)*	0.1354	(0.0192)*
obs. number	10 811		10 811		10 811		10 811	

(*) Significant at 95% level.
(**) Significant at 90% level